

# PRÉVISION DE LA CONSOMMATION ÉLECTRIQUE PAR MÉLANGE DE PRÉDICTEURS : TRAVAIL SUR LE JEU D'EXPERTS

Pierre Gaillard <sup>1</sup> & Yannig Goude <sup>2</sup>

<sup>1</sup> *pierre@gaillard.me* <sup>2</sup> *yannig.goude@edf.fr*

**Résumé.** Dans cette étude empirique, nous nous intéressons à la prévision à court terme (horizon journalier) de la consommation électrique française. L'électricité se stockant difficilement, il s'agit d'un enjeu important pour une entreprise comme EDF qui doit équilibrer la production avec la demande d'électricité. EDF R&D a ainsi développé ces dernières années plusieurs modèles de prévisions basés sur des approches statistiques très différentes (analyse de données fonctionnelles, de séries temporelles, régression semi-paramétrique). Nous explorons ici comment utiliser ces modèles afin d'améliorer la qualité de prévision. Nous commençons par agréger les prévisions initialement à disposition. Afin de réduire encore l'erreur de prévision du mélange, nous proposons ensuite différentes approches pour construire de nouvelles prévisions à partir des méthodes initiales.

**Mots-clés.** Prévision de séries temporelles, apprentissage séquentiel, suites individuelles, mélange d'experts

**Abstract.** We are interested in this empirical study in daily forecasting of French electricity load consumption. Since energy storage is a difficult challenge, it is a matter of great importance for companies such as EDF to maintain the equilibrium between production and demand. Short term electricity forecasting has thus been studied for years at EDF R&D and different forecasting models have been developed from various field of statistical or machine learning theory (functional data analysis, time series, semi-parametric regression). We investigate how we can use this initial forecasting models to improve the accuracy of the predictions in different ways. First, we show how combining the original set of forecasts can lead to a significant improvement of the forecasts. Second, we explore how to build from these models various and heterogeneous forecasts and analyze how we can aggregate these forecasts to get even better predictions.

**Keywords.** Time series forecasting, individual sequences, load forecasting, mixture of experts

## 1 Introduction

La prévision de la consommation électrique représente un enjeu important pour une entreprise comme EDF afin d'assurer l'équilibre entre la production et la demande. EDF

R&D a ainsi développé plusieurs modèles de prévision atteignant environ 1.4% d'erreur MAPE à un horizon journalier. Le paysage électrique français est cependant en constante évolution, que ce soit au niveau des usages (voitures électriques, ampoules basses consommation,...), de la production (nucléaire, énergies renouvelables) ou du marché européen avec l'ouverture à la concurrence. Les modèles historiques se retrouvent ainsi régulièrement remis en question. Plutôt que de sélectionner un de ces modèles, on se propose ici d'améliorer la qualité de prévision en suivant une méthodologie en deux étapes. Tout d'abord, on se propose d'agréger de façon robuste et séquentielle trois modèles de prévision existants. Ensuite, nous cherchons à enrichir notre ensemble de prévisions à inclure dans le mélange en créant de nouvelles prévisions suffisamment variées à partir des méthodes de prévisions initiales.

EDF R&D a développé un nombre considérable de méthodes de prévisions de la consommation. Parmi celles-ci nous nous concentrons sur trois méthodes que nous présentons ci après. Nous les avons choisis non seulement parce qu'elles ont une bonne qualité de prévision; mais aussi car elles reposent sur des modèles statistiques très différents, ce qui induit une bonne variété dans leurs prévisions. Nous espérons que les mélanger permettra de tirer avantage de leurs comportements différents. Le premier modèle est un modèle semi-paramétrique effectuant une régression pénalisée sur une base de splines. Il a été introduit par Wood (2006) et nous le nommerons modèle additif généralisé (GAM). Le deuxième est basé sur une régression linéaire dans un espace fonctionnel (CLR). Il a été introduit et appliqué à la consommation électrique par Cho et al. (2013). Le troisième et dernier modèle (KWF) est détaillé dans Antoniadis et al. (2010) et associe un clustering de données fonctionnels avec la détection de similarités dans une base d'ondelettes.

## 2 Prévision séquentielle à l'aide d'experts

Les modèles stochastiques que nous voulons agréger reposent sur des hypothèses très différentes. Nous adoptons donc la théorie des suites individuelles, détaillée dans l'excellent livre de Cesa-Bianchi et Lugosi (2006), qui propose un cadre robuste permettant de les associer sous un même toit.

Nous supposons que nous disposons d'une suite d'observation  $y_1, \dots, y_T \in [0, B]$  bornées à prévoir de manière séquentielle. Dans notre application, il s'agit de la consommation électrique française. Avant chaque instance  $t$ , un ensemble fini de  $K$  experts (ici les trois méthodes de prévision initiales) nous proposent des prévisions  $\mathbf{x}_t = (x_{1,t}, \dots, x_{K,t}) \in [0, B]$  de la prochaine observation  $y_t$ . Le statisticien doit alors proposer sa propre prévision  $\hat{y}_t$  de  $y_t$  à partir de sa connaissance des observations passées  $y_1, \dots, y_{t-1}$ , ainsi que des prévisions des experts présentes et passées  $\mathbf{x}_1, \dots, \mathbf{x}_t$ . Dans ce but, nous supposons qu'il forme un vecteur de mélange  $\mathbf{p}_t = (p_{1,t}, \dots, p_{K,t})$  et prévoit la combinaison linéaire  $\hat{y}_t = \mathbf{p}_t \cdot \mathbf{x}_t$  des prévisions des experts.

Nous évaluons la performance des experts et du statisticien par leurs erreurs quadratiques moyennes. Le but de statisticien étant de minimiser sa perte moyenne qui peut se reformuler en la somme des deux termes

$$\frac{1}{T} \sum_{t=1}^T (y_t - \widehat{y}_t)^2 \triangleq \inf_{\mathbf{q} \in S} \left\{ \frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{q} \cdot \mathbf{x}_t)^2 \right\} + R_T, \quad (1)$$

où  $S$  est un ensemble fermé et borné de  $\mathbb{R}^d$ . Cette écriture exprime le compromis connu en statistiques entre erreur d'approximation (1er terme) et erreur d'estimation (2ème terme).

Le premier terme représente la performance atteignable par le meilleur vecteur de mélange constant (oracle) que le statisticien aurait pu choisir. Plus notre ensemble d'experts inclus dans le mélange sera hétérogène plus l'erreur d'approximation sera faible mais plus le statisticien aura du mal à l'atteindre.

Le deuxième terme représente la difficulté du statisticien à estimer l'oracle de façon séquentielle. De nombreuses méthodes d'agrégations permettent d'assurer un regret moyen  $R_T$  tendant vers 0, et ce de façon déterministe, c'est à dire qu'elles que soient les prévisions des experts  $\mathbf{x}_t$  et les observations  $y_t$ .

Nous présentons maintenant rapidement les deux stratégies d'agrégations étudiées. La première est la regression linéaire Ridge (voir Azoury et Warmuth, 2000), qui choisit le vecteur de poids minimisant un critère des moindres carrés pénalisé. Le paramètre de lissage  $\lambda$  est optimisé de façon séquentielle sur une grille (voir Devaine et al. 2014). La seconde stratégie est l'algorithme MLpol proposé par Gaillard et al (2014) qui attribue des poids à partir de potentiels polynomiaux en considérant un paramètre d'apprentissage différent par expert. Les poids formés par cette stratégies sont convexes : ils sont positifs et somment à 1. Cette stratégie propose de plus une calibration théorique efficace des paramètres d'apprentissage que nous utilisons.

### 3 Expériences

Nous considérons une version mise à jour du jeu de données analysé dans Devaine et al (2013). Il contient à un pas 30 min la consommation électrique du portefeuille EDF français du 1er janvier 2008 au 15 juin 2012, ainsi que plusieurs variables exogènes (météo, type de jour, ...). Les jours atypiques (jours fériés  $\pm 2$ , changement d'heures) sont retirés. L'objectif est de prévoir chaque jour à midi la consommation de la journée suivante (i.e. les 48 prochaines demi-heures). Nous utilisons les jours du 1er janvier 2008 au 31 août 2011 pour entrainer les experts, puis nous évaluons leurs performances entre le 1er septembre 2011 et le 15 juin 2012.

Le Tableau 1 retranscrit les performances du jeu d'experts ainsi que des méthodes d'agrégation. Le simple fait de mélanger les prédicteurs permet une diminution d'environ 15% de l'erreur de prévision.

Oracles et stratégies d'agrégation	RMSE (MW)	MAPE (%)
Meilleur expert	744	1.29
Meilleure combinaison convexe	629	1.06
Meilleure combinaison linéaire	629	1.06
MLpol	626	1.05
Ridge	638	1.06

Table 1: Erreurs des oracles et stratégies d'agrégation pour l'ensemble d'experts  $E_0$ : GAM, CLR, and KWF.

Nous enrichissons maintenant l'ensemble d'experts afin d'améliorer la performance du mélange. Nous avons construit de nouveaux experts via quatre méthodes que nous détaillons maintenant rapidement :

- **Bagging.** La première méthode est le bagging, une méthode d'apprentissage statistique proposée par Breiman (1996). Celle-ci consiste à entraîner la méthode statistique non pas sur l'ensemble du jeu d'entraînement mais sur différents échantillons bootstrap tirés aléatoirement. Nous construisons ainsi 15 nouveaux experts pour chacune de nos trois méthodes initiales et nous notons  $E_1$  cet ensemble de 45 experts.
- **Spécialisation.** La deuxième est la spécialisation, et consiste à pondérer différemment les jours lors de l'entraînement en fonction des variables exogènes de façon à donner plus d'importance à certains scénarios météorologiques (canicule, variation brutale de température, ...) Nous imaginons 8 scénarios que nous entraînons sur nos trois méthodes initiales, ce qui nous amène à un nouvel ensemble  $E_2$  de 24 experts.
- **Deux échelles temporelles.** Cette méthode est détaillée dans Nedellec et al. (2014). Elle consiste à entraîner nos méthodes sur deux échelles temporelles. On commence par modéliser la consommation au niveau moyen terme (i.e., sans utiliser les données de consommation récentes); le modèle est ensuite recalculé au niveau court terme sur les 30 derniers jours. On crée ainsi un nouvel expert  $E_3$  à partir de la méthode Gam.
- **Boosting.** Cette dernière stratégie pour créer de nouveaux experts est inspirée du boosting. Elle consiste à ajouter de nouveaux experts qui vont tenter de corriger les erreurs des experts initiaux. Pour cela, plutôt qu'entraîner une méthode à prévoir directement la consommation d'électricité  $y_t$ , on l'entraîne à prévoir  $(y_t - \gamma x_{k,t}) / (1 -$

$\gamma$ ), avec  $\gamma \in (0, 1)$ , de façon à compenser en partie les erreurs effectuées par l'expert  $k$ . Nous créons ainsi un ensemble  $E_4$  de 45 nouveaux experts en variant la valeur de  $\gamma \in \{0.5, \dots, 0.9\}$  et les méthodes utilisées.

La Figure 1 représente les pertes subies par les stratégies d'agrégations en fonction des ensembles d'experts inclus dans le mélange. Nous voyons que mélanger et enrichir l'ensemble d'expert a permis une réduction d'environ 25% de l'erreur, une partie en mélangeant simplement les experts, l'autre en créant de nouveaux experts. Nous avons de plus remarqué que le mélange est ici assez peu sensible au sur-apprentissage du à un trop grand nombre d'experts.

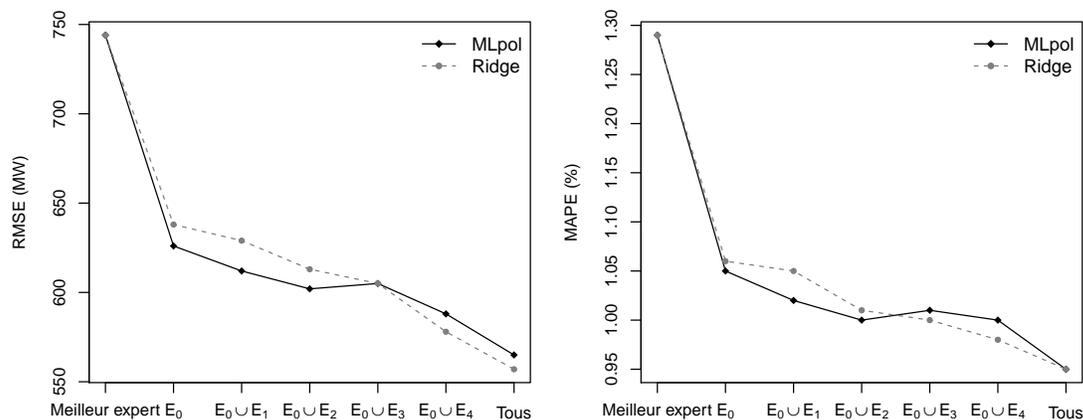


Figure 1: Comparaison de la performance de la meilleure méthode initiale dans  $E_0$ , avec celle obtenue en mélangeant les experts de  $E_0$ ,  $E_0 \cup E_1$ ,  $\dots$ ,  $E_0 \cup E_4$  en utilisant les stratégies MLprod et Ridge, et en mélangeant tous les experts à disposition (Tous).

## Bibliographie

- [1] Antoniadis, A., Brossat, X., Cugliari et J., Poggi, J.(2010), *Clustering functional data using wavelets*, In: Proceedings of the Nineteenth International Conference on Computational Statistics (COMPSTAT).
- [2] Azoury, K.S. et Warmuth, M.K. (2000), *Relative loss bounds for on-line density estimation with the exponential family of distributions*, Machine Learning 43(3)
- [3] Breiman, L. (1996) *Bagging predictor*, Machine Learning 24(2)
- [4] Cesa-Bianchi, N. et Lugosi, G. (2006), *Prediction, Learning and Games*, Cambridge University Press
- [5] Cho, H., Goude, Y., Brossat et X., Yao, Q. (2013), *Modeling and forecasting daily electricity load curves: a hybrid approach*, Journal of the American Statistical Association

- [6] Devaine, M., Gaillard, P., Goude, Y. et Stoltz, G. (2013), *Forecasting the electricity consumption by aggregating specialized experts*, Machine Learning 90(2)
- [7] Gaillard, P. et Goude, Y. (2013), *Forecasting the electricity consumption by aggregating experts*, submitted
- [8] Gaillard, P., Stoltz, G., et van Erven, T. (2014), *A Second-order Bound with Excess Losses*, submitted
- [9] Nedellec, R., Cugliari, J., Goude, Y. (2013), *Gefcom2012: Electric load forecasting and backcasting with semi-parametric models*, International Journal of forecasting, **in Press**
- [10] Wood, S. (2006), *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC